ENTRAPMENT, TEMPTATION, AND VIRTUE TESTING

1. INTRODUCTION

Cases of entrapment involve a party that intends to entrap, whom we call the 'agent', and a party that is entrapped, whom we call the 'target'. Let the terms 'party', 'agent', and 'target' encompass both individuals and groups.

In this article we focus on the ethics of entrapment, temptation, and virtue testing. In particular, we look at the question of the moral permissibility or impermissibility of these three actions. In Section 2, we adopt and summarize the account we gave in (Hill, McLeod, and Tanyi, 2018) of the concept of entrapment. We then contrast entrapment with intentional temptation and with virtue testing. We explain how an agent can intentionally present a person with an opportunity to act in a way believed to be impermissible, but *without* intending that the person take up the opportunity to act in this way. Procurement involves both intentions. While entrapment, temptation, and virtue testing all involve the former intention, only entrapment necessarily involves procurement.

In Section 3, we discuss a moral objection to entrapment that we call 'the objection from temptation'. We focus on a version of this objection proposed by Hughes (2004). On his account, the wrong of entrapment resides mainly in the fact that in tempting the target, the agent compromises or undermines the autonomy that is necessary for the target properly to be held culpable for the entrapped act. After explaining why we regard Hughes's version of the objection as unsound, we strip his version of the objection down into a simpler, more plausible, argument from temptation. According to it, the agent intends that the target should fall to the temptation to perform an act considered to be impermissible by the agent; it is impermissible to intend that an act that one considers impermissible be performed (whether by oneself or another); thus, entrapment is impermissible.

We respond first of all that not every case of entrapment is a case of temptation. While there are cases of entrapment in which the agent succeeds in tempting the target, there are also

cases in which the agent does not intend to tempt, and cases in which, while the target performs the impermissible act, it would be incorrect to describe the target as having been tempted (as against merely motivated) to perform it. This significantly limits the scope of the objection from temptation, both in the form given by Hughes (2004) and in the simpler form arrived at in the course of our critique of his position.

Removal of the appeal to temptation leaves us with an even sparser argument against entrapment, which we call 'the objection from intention'. For those acts of entrapment that are of present interest (see Section 2) it locates wrong-doing simply in the agent's intention that the target act in a manner considered impermissible by the agent (regardless of whether the target is, along the way, tempted): the impermissibility of the target's entrapped act spreads to the agent's act of entrapment. We dub this ethical principle—the principle that if one considers that an act is impermissible then it is impermissible for one to intend that it be performed (whether by oneself or another)—'the Purist Principle'. In Section 4, we discuss and reject this principle.

In Section 5, we present a new moral objection to entrapment that we call 'the objection from moral alliance'. According to it, acts of entrapment are typically morally faulty at least partly because an agent that entraps a target thereby becomes morally allied (in a manner that we explain) with the target's impermissible act in procuring it.

In Section 6, we address the morality of virtue testing and intentional temptation: since these are exempt from the objection from moral alliance, they are easier for an agent to justify. Virtue testing with the intention that the target pass the test is generally permissible, although the risks and benefits involved can change the picture. The permissibility of virtue testing with the opposite intention, and that of intentional temptation, are generally subject to two conditions drawn from our discussion of the Purist Principle: first, that the target was likely to have performed the entrapped act (or something else as bad or worse) later anyway, and,

secondly, that the entrapped act (or the equally bad or worse act) will not now be performed, or that its harmful consequences will be nullified. Nevertheless, we argue that further considerations, concerning risks, harms, benefits, and motives, can justify exceptions to these conditions in extreme cases. Section 7 is a concluding summary.

2. CONCEPTUAL BACKGROUND

This section builds on material that we advanced in (Hill, McLeod, and Tanyi, 2018). It then contrasts entrapment with intentional temptation and with virtue testing. Initially, we draw two distinctions, which cut across each other, concerning acts of entrapment. The first concerns the status of the agent and the second concerns the act that the target performs and that the agent has procured.

Our first distinction concerns the agent's status. *Legal entrapment* occurs when the agent is a law-enforcement officer, acting (in accordance with the law or otherwise) in an official capacity as a law-enforcement officer, or when the agent is acting on behalf of a law-enforcement officer, as the officer's deputy. An example of this would be an undercover police officer's asking a suspected drug pusher to sell drugs in order to arrest the suspect *in flagrante*. When, on the other hand, the agent is neither a law-enforcement officer acting in that capacity nor the deputy of such an officer, acting as a deputy of the officer, we have *civil* entrapment. An example of this would be a vigilante's posing as a child and suggesting, in order to secure an arrest, to a suspected child molester that he come and molest her.

Our second distinction concerns the procured act: we distinguish between acts that are of a criminal type, as in the above examples, and those that are not. For an example of the latter case, an investigative journalist might entrap a politician into performing a morally compromising act that is not a crime, in order that the journalist might expose the politician for having performed the act. Or, to take an example involving law-enforcement agents: suppose

that the security service, in their capacity as law-enforcement agents, entrap an enemy spy using diplomatic cover into making an embarrassing boast that will lead his superiors to recall him thus preventing him from performing the illegal act of espionage. When the act is not criminal but is morally compromising in some way (whether by being immoral, embarrassing, or socially frowned upon in some way), we are dealing with *moral entrapment* (using the word 'moral' in a wide sense). When the act is of a criminal type, we have *criminal entrapment*.

Thus, four types of entrapment can be distinguished: legal criminal entrapment, civil criminal entrapment, civil moral entrapment, and legal moral entrapment. In what follows, we restrict our concern to cases of entrapment in which the agent intends to entrap the target into performing an action that the agent considers to be morally impermissible. Such cases, we take it, may be drawn from all four of these types of entrapment.

In (Hill, McLeod, and Tanyi, 2018) we argued that entrapment occurred whenever:[1]

(i) an agent plans that a particular act be committed;

(ii) the planned act is of a type that is criminal, immoral, embarrassing, or socially frowned upon (measurable in part by the extent to which the target would probably not like the act to be exposed to colleagues, an employer, friends, family, or the public);

(iii) the agent procures the act (by solicitation, persuasion, or incitement);

(iv) the agent intends that the act should, in principle, be traceable to the person performing it (the target) either by being detectable (by a party other than the target) or via testimony (including the target's confession), that is, by evidence that would link the target to the act;

[1] We provided in (Hill, McLeod, and Tanyi, 2018) a *philosophical* account of entrapment. We do not assert that this account will precisely fit the specialized *legal* use of the term.

(v) in procuring the act, the agent intends to be enabled, or intends that a third party be enabled, to prosecute or to expose the target for having performed the act.

While it is not a consequence of these conditions that whenever entrapment happens the target performs an act that *the agent considers* impermissible, we restrict our discussion of the morality of entrapment (and of intentional temptation and virtue testing), in subsequent sections, to such cases.[2]

In (Hill, McLeod, and Tanyi, 2018) we provided an extended discussion and defence of these conditions, which built on the work of Stitt and James (1984), Gerald Dworkin (1985) and Ho (2011).[3] The wording of (i) here differs a little from the wording that we used in (Hill, McLeod, and Tanyi, 2018). This is to make it clear that it is not necessary for the agent to intend, *de re*, of a particular target that the target perform the act.[4] Miller and Blacker (2005: 102–109) distinguish between *targeted* action by an agent, aimed at some target in particular, and *random* action, aimed at no-one in particular. For example, if an agent attempts to procure an impermissible act from a person that the agent already intends should perform it, then the agent's act of attempted procurement is a targeted one. If, on the other hand, the agent arranges a posting to a billboard encouraging those that happen to pass by to perform an impermissible act, then the agent's act of attempted procurement is random. The terms 'entrapment', 'intentional temptation', and 'virtue testing' as used here encompass both targeted and random acts.

---

[2] This is because the particular moral problems on which we wish to focus are raised by these cases, but not by cases such as when a journalist entraps a celebrity into doing something immoral that the journalist himself or herself does not consider immoral. This is not to deny that these other cases themselves raise moral problems.

[3] On condition (i), cf. (Feinberg, 2003: 73–4).

[4] Thanks to Tom Brown for his help here.

Condition (v) includes, we take it, blackmail cases in which the agent intends not that the target will be prosecuted or exposed, but that the target will be placed under *threat of prosecution or exposure*.

Condition (iii), the procurement condition, amends that of Dworkin (1985: 21), replacing his 'enticement', which includes non-communicative acts, with 'incitement', which does not.[5] Given that the account of procurement here is a communicative one, it encompasses what Feinberg (2003: 58) calls 'goading'. In order to understand the position, it is important to note that (iii) implies that the target actually does perform, rather than merely attempt, the act planned under condition (i).[6]

The definition of entrapment used here employs a very specific understanding of procurement. Procurement is understood to involve the agent's having an intentional influence, via directly related communicative acts, on the target's will. If all that the agent does is intentionally present the target with the opportunity to perform the act, then this does not amount, on this account, to entrapment. Also, what Feinberg (2003: 74) describes as persuasion 'by noncoercive manipulation' counts, on this account, as procurement, though we also count as procurement other cases that Feinberg does not count as procurement.

If one procures an offence one causes it, albeit indirectly, to occur. Not all cases of causation are cases of procurement: given this account's appeal to communicative acts, if *A* causes *B* to perform an action by administering a drug, then *A* has not procured that action from *B*. By contrast, if *A* threatens *B* by saying that *A* will shoot *B*, or expose *B*'s darkest secrets,

[5] We understand that different legal jurisdictions attach different levels of significance to the notion of procurement and that the ensuing, purely philosophical account, of procurement that we provide differs from those in common legal currency. We believe, nevertheless, that its doing so is subsequently useful when discussing types of act that are, no matter what we might call them, importantly ethically different

[6] We return to this distinction towards the end of the article.

unless *B* performs a certain action, then *A* has procured that action, provided that *B* does it under the influence of *A*'s threat. Likewise, if *A* influences *B*'s will in a similar fashion, but through a bribe, then this counts as procurement. Procurement always goes by appeal to the will, even if the target acts irrationally in performing the procured action. Procurement need not involve an inducement of pleasure to be had or pain to be avoided. One can procure an action merely by recommending it or requesting it. The phrase 'recommending it or requesting it' must be construed so as to include heavy hints and the use of implicature to recommend (as in 'It would be good if our competitor could be made to disappear'). Recommendations and requests can also be communicated via such methods as gestures, drawings, mime, and sign-language.

We take it that every act of procurement involves *intentional* causation (but not *vice versa*). If one unintentionally causes someone to perform a wrongful act, then that is not a case of procurement. (It does not follow that in procuring one always intends that one *procure*: many people engaged in procurement are perhaps quite unaware that there is such a thing as procurement, and are fully concentrated on the causation.)

Let us apply this account of procurement, which appeals to the influence upon the target's will of an act, or series of acts, of communication by the agent, to a couple of examples. If the agent leaves a wallet lying in the street in order to trap the target, then the agent has not procured the act of taking the wallet unless the agent in some way recommends the action to, or requests the action of, the target. It is clear that if an agent actually asks a suspected drug dealer to sell drugs then that is a case of procurement. By contrast, if the agent just hangs around hoping that the drug dealer will approach and offer the drugs for sale, then that is not procurement, even if the agent uses a disguise to resemble a potential client.

Let us now turn to the notion of *temptation*. Whether a person is tempted primarily concerns their emotions, rather than their actions. In our view, when a target is tempted, the

target experiences an urge to perform an act, while also being to some degree internally conflicted about that urge (cf. Hughes, 2004; 2006a; 2006b). If a person feels tempted to perform an act and, as a direct result, attempts to perform it, then that person has succumbed to temptation (even if the attempt to perform the act is unsuccessful).

We distinguish between *intentional temptation*, in which the target is tempted (even if they do not succumb) by an agent intending to tempt them, and merely *situational temptation*, in which someone is in circumstances that they interpret as providing an alluring opportunity. A situation may also be said to be tempting when it would tend towards the situational tempting of persons put into it. Whereas the situational temptation of an agent is an occurrent matter, that a situation is tempting in the sense just mentioned is a dispositional fact about it. Except in passing, we are concerned only with intentional temptation.

Even if it is supposed, against our own account in Section 3, that entrapment always involves temptation, there is a difference between being entrapped into the performance of an act and performing it as a direct result of mere temptation. In entrapment the agent seeks to bring about the performance of the act by procurement, that is, by means of certain directly related communicative acts. For example, the agent might ask the target 'Do you have any drugs for sale?'. Or the agent might use the imperative form 'Please sell me some drugs!' or might make a statement with an obvious implication such as 'I'd like to buy some drugs'. The meaning of 'directly related' should be clear in view of the contrast that we now draw.

For there to be temptation, however, there need not be any procurement, or attempted procurement, of the impermissible action. If the agent seeks to bring about the performance of the act by means of non-communicative wiles, or by indirectly related communicative acts, then we do not have a case of procurement, and therefore we do not have a case of entrapment, even if the agent succeeds in tempting the target into the performance of the act. For example, suppose that the agent is posing as a senior citizen with the intention of luring out a mugger.

The agent might spy a target, and then try to tempt the target to mug by non-communicative schemes, such as adopting a doddery walk, 'accidentally' dropping some cash, adopting the air of one lost etc. Or the agent might even engage in various (indirect) communicative acts, such as saying to the target 'I'm lost and am looking for the bank; can you help?'. Here the agent neither requests nor orders the target to mug; rather, the agent's communicative acts are intended to create in the target's mind certain beliefs that are meant to elicit an intention, on the target's part, to mug the agent (such as the belief that the agent is vulnerable and has money on their person). Although temptation need not feature procurement, and what we call 'mere temptation' does not feature procurement, nevertheless, it may do so.

Virtue testing is, on our account, distinct from intentional temptation. In a case of virtue testing, the agent presents the target with the opportunity for wrong-doing, but without necessarily having the intention of bringing it about that the target perform, or be tempted to perform, the wrongful act, and without necessarily procuring the wrong action. For example, suppose that an agent disagrees with a colleague over whether another colleague, Kim, is a thief. Suppose further that the agent is confident not only that Kim is not a thief, but also that Kim will not readily experience any urges to steal when presented with the opportunity for theft. The agent might say to the colleague 'I can prove to you that Kim is not a thief; let's leave some money lying out as a test, and you'll see that Kim won't take it; in fact, I'm certain that Kim won't even be tempted'. In this scenario, the agent does not intend that Kim take the money or feel the urge to take the money; in fact, the agent has the opposite intentions. Nevertheless, the agent does intend to provide Kim with the opportunity for wrong-doing.

Moreover, whenever we trust someone to do something, knowing that they might breach our trust, we intentionally present them with an opportunity to do wrong. Doing so is, in itself, innocuous. To trust a trustworthy person is not to tempt them (and is certainly not to entrap them), even though it may be a test for them.

Virtue testing, as we define it, occurs when an agent intentionally presents a target with an opportunity to perform an act that the agent considers impermissible, in order to discover (or demonstrate) whether the target is disposed to perform it. When an agent intentionally tempts a target, the agent presents the target with this opportunity having, furthermore, and among other things, the intention that the target will experience an urge to perform the impermissible act. Virtue testing is consistent with, but does not entail, intentional temptation. For two parties, *A* and *B*, if *A* tests *B*'s virtue then it does not follow that *A* tempts *B*, or even that *A* intends to tempt *B*, though it is possible for a case of virtue testing to be also a case of temptation, and, indeed, a case of entrapment too.

To sum up, in the scenarios of virtue testing, of intentional temptation, and of entrapment that are of present interest, a common element is the intentional presentation to the target, by the agent, of the opportunity to perform an act that the agent considers impermissible. That is, the agent intentionally puts the target into a situation that the target might interpret as providing an opportunity for wrongdoing. While in cases of entrapment we have procurement (with or without intentional temptation), in cases of intentional temptation, although we have temptation, we need not have procurement, and in cases of virtue testing we need have neither procurement nor intentional temptation.[7] For the time being, we focus on entrapment. We return to virtue testing and intentional temptation near the end of the article.

[7] Contrast the taxonomies of, for example, (Dworkin, 1985) and (Miller and Blackler, 2005: 102–109). On the distinction between virtue testing and tempting, compare (Hughes, 2004: 56).

3. THE OBJECTION FROM TEMPTATION

The definition of entrapment we employed was neutral about whether the agent, in entrapping, acted permissibly; we take this to be a desirable feature of a definition of entrapment. Over the next three sections, we are primarily concerned with the question of permissibility.

The basis of the objection from temptation is that there is, in relevant circumstances at least, something wrong with intentional temptation. According to the objection, entrapment, at least of a given sort, involves intentional temptation, and is wrong for that reason, or else, if it does not involve intentional temptation itself, it is wrong because it shares the feature that makes intentional temptation wrong.

A *target-centred* version of the objection from temptation says that entrapment is wrong because of the effects that the agent's act of entrapment has on the target's mental states or agential capacities. An *agent-centred* version, which we discuss in more detail in Section 4, says that entrapment is wrong because the entrapping agent has the impermissible intention that the target succumb to the temptation to perform an impermissible act. It is the focus on the acts or intentions of the agent that makes a version of the objection from temptation agent-centred. The specific agent-centred version of the objection that interests us works with the general principle that if an act is impermissible then it is thereby impermissible that one should intend that it be performed. We dub this principle 'the Purist Principle'.[8]

The first version of the objection from temptation that we wish to consider is from (Hughes, 2004). He argues that legal criminal entrapment, when it is without 'probable cause', is wrong for the first of the reasons just mentioned: that is, because it involves intentional

[8] The Purist Principle is closely related to, but broader than, the principle that 'formal co-operation' with something impermissible is impermissible. On the notion of formal co-operation see (Smith, 1995).

temptation.[9] According to (Hughes, 2004: 45, cf. 55), this sort of 'entrapment […] illegitimately violates' the target's moral autonomy. For Hughes (2004: 54–55), an agent that entraps in this way acts wrongly because that agent deliberately subjects the target to temptation. Whenever an agent tempts a target, then, in turn, the target's 'will is locked in an internal conflict, and this compromises or sometimes even undermines [the target's] autonomy' (Hughes, 2004: 53). In cases of legal criminal entrapment without probable cause, it is this sort of manipulation of the target's will that makes the entrapping agent's act wrong (Hughes, 2004: 56). As Hughes (2006a: 224, his italics, cf. 2006b: 355) summarizes his position:

> entrapment, like temptation in general, compromises or even undermines the autonomy of those entrapped. Since autonomy is required for reasonable ascriptions of moral and legal responsibility, the *moral* problem with entrapment is that it defeats an essential condition of criminal liability in the process of detecting and preventing crime.

The position that Hughes endorses is as follows. Entrapment must involve the intentional temptation of the target by the agent. This temptation compromises or undermines the target's autonomy, in a manner that is incompatible with the target's culpability for the entrapped act. The view would be that these prerequisites hold for every entrapment scenario, from any of the four types discussed in Section 2. Thus, given the absence of culpability, no punishment, opprobrium, or ostracism of the target would be justifiable. Importantly, the notion of culpability in play here is a broadly moral notion distinct from, albeit closely related to, that

[9] Park (1976: 270–73) and Ho (2011: 72–73), among many authors, take it that every act of entrapment involves intentional temptation, though they do not explicitly offer this as a moral objection to entrapment.

of criminal responsibility.[10] Accordingly, a generalized argument from temptation that is recoverable from Hughes's work goes as follows:

1.	Whenever an agent, *A*, entraps a target, *B*, *A* intentionally tempts *B* to perform the entrapped act, for which *A* intends that *B* will be held culpable.	Premise
2.	When *B* is so tempted, *B*'s autonomy is either compromised or undermined.	Premise
3.	One is culpable for an act that one has performed only if, in performing the act, one's autonomy is not compromised or undermined.	Premise
4.	If a person is not culpable for an act, then it is wrong for that person to be held culpable for it.	Premise
5.	If an act is wrong, then it is wrong to act with the intention that it be performed (whether by oneself or by another).	Premise
6.	Whenever an agent, *A*, entraps a target, *B*, then *B* is not culpable for the entrapped act.	From 1, 2, 3
7.	Whenever an agent, *A*, entraps a target, *B*, then it is wrong for *B* to be held culpable for the entrapped act.	From 4, 6
8.	Whenever an agent, *A*, entraps a target, *B*, then *A* acts wrongly.	From 1, 5, 7

[10] Cf., on the distinction, (Feinberg, 2003: 67–8, 75): it includes blameworthiness for non-criminal acts, for non-harmful acts (e.g., offensive or rude acts), and for acts for which, while either harmful or of a criminal type, the target is not ultimately held criminally or civilly liable by a court (e.g., because the target is deemed by the court to have been entrapped).

Before we return to the specifics of Hughes's own position, let us evaluate some aspects of this more general argument. Performing this task serves two roles: it figures, via evaluation of material common to both, in our evaluation of Hughes's own, more restricted, version of the objection from temptation and it facilitates the transition into the next phase of our own argument.

The argument is valid. Its conclusion follows because, according to Premise 1, an entrapping agent intends that the target be held culpable for the entrapped act; this is, according to 7, the intention, in respect of a wrongful act (i.e., the holding culpable), that it be performed and, according to 5, intending, in respect of a wrongful act, that it be performed is itself wrongful. (We have been careful to word this so that opacity is not introduced where we do not intend it.) Setting aside Premise 1, for now, we note that Premises 2 and 3 are controversial. With respect to Premise 2: even if there are temptations that are irresistible, and thus that compel, to tempt is not, *per se*, to compel, and it might be plausibly maintained that resistible temptations do not necessarily compromise or undermine the target's autonomy. With respect to Premise 3, since compromise to autonomy is, like culpability, a matter of degree, Premise 3 is too broadly worded. Even if, contrary to our view, entrapment always involves temptation, and if it thereby involves compromise to *B*'s autonomy, compromise is not all or nothing. To the extent that *B*'s autonomy was not compromised, and in the absence of any other factor that exonerates *B*, *B* remains to that extent culpable (cf. Fletcher, 1978: 542; Altman and Lee, 1983: 59; Feinberg, 2003: 58; Dillof, 2004: 895). Given the problematic nature of Premises 2 and 3 of the above argument, we now shift our focus to an agent-centred version of the objection from temptation that dispenses with these premises. The principle put forward in Premise 5, which we called earlier 'the Purist Principle', features in this argument, and occupies an important role in the section after that.

The new, agent-centred, argument that we wish to consider may be summarized as follows:

1.	If an act is impermissible, then it is impermissible to act with the intention that it be performed (whether by oneself or by another).	Premise
2.	In entrapment, the agent intends, in respect of an impermissible act and a target, that the target should fall to the temptation to perform that impermissible act.	Premise
3.	Entrapment is impermissible.	From 1, 2

Premise 1 of the argument in the previous section overlooked the possibility of a case of entrapment for purposes of blackmail, in which the agent did not intend that the target be held culpable; Premise 2 here does not have that flaw. A more serious problem remains, however, namely that not all cases of entrapment actually involve intentional temptation. In the definition of entrapment that we set out in Section 2, the notion of temptation played no role. On our account, in entrapment scenarios the agent's crucial intention is that the target perform the act that the agent is trying to procure. What goes on in the target's mind, other than the inclination of the target's will towards performance of that act, the formation of the intention to perform it, and the execution of that intention, need not be of concern to the agent.[11] Indeed, an agent that agreed with Hughes's views on the relationship between temptation and culpability might intend to procure the act just so that the target could properly be held culpable

[11] While, as stated, we take procurement to be a specific form of intentional inducement, Miller and Blackler (2005: 102–105), are right, we think, to talk about *inducement* into the entrapped act, rather than the *temptation* to perform it, which they do not even mention. Miller, Blackler, and Alexandra (2006: 264) depict inducement as something that a target may 'resist', but they do not explicitly link it to temptation.

for it, and might intend that the target merely perform it, i.e. perform it with no ambivalence or mental anguish along the way, rather than perform it as a direct result of having been tempted and succumbed. Such an agent wishes the target to be *motivated* to perform the act, but is at best indifferent about whether the target is *tempted* to perform it.[12] Even if entrapment commonly involves temptation, it does not, or at least need not, always do so. If acts of entrapment are wrong, and are so for a reason related to temptation, then at best this can be because they are importantly morally analogous to acts of intentional temptation; it cannot be because every act of entrapment essentially involves an act of intentional temptation.

4. THE PURIST PRINCIPLE AND THE OBJECTION FROM INTENTION

If we strip away from the agent-centred version of the objection from temptation the appeal to temptation, we are left with the following valid argument, which we call 'the objection from intention'. Its first premise is the Purist Principle that we saw at work in each of the two preceding arguments.

1. If an act is impermissible, then it is impermissible to act with the intention that it be performed (whether by oneself or by another).[13] Premise

[12] A person can be motivated, for example, to eat a healthy diet or to take exercise, without being tempted to do such things. To be tempted is to have a kind of motivation, with a specific phenomenology; motivation is a broader category.

[13] This premise could be restricted by adding that the agent must believe, or know, that the act is impermissible. This will not matter for what follows.

2.	In the cases of entrapment that are of interest to us (see Section 2), the agent has such an intention.	Premise
3.	Entrapment is, in such cases, impermissible.	From 1, 2

Premise 2 of this argument is true as a matter of stipulation. It is Premise 1, the Purist Principle, that is of philosophical interest.

Like intentional temptation towards wrongdoing, entrapment breaches the Purist Principle, expressed in Premise 1, that *if a certain action is impermissible then it is also impermissible to intend that it be performed*. To be clear, this principle does not require that the agent's motivating thought be that the target perform an impermissible action. Rather, it suffices that the agent intends, with respect to an impermissible action, that the target perform it.

To see the initial appeal of the Purist Principle, take the following example. The Mafia boss wants to keep his own hands 'clean', and so manipulates the easily led tyro into killing a rival that is threatening the boss's interests. This manipulation need not take the form of full-blown determinism; it could be that the boss is a shrewd judge of character and can predict how the tyro is likely to react in different situations. The boss's predictions need not even be more than guesses; if the mafia boss intends, while recognizing that this intention may not be fulfilled, that the tyro kill the rival this seems enough to establish that the Mafia boss is morally at fault for having such intentions.

Although the Purist Principle might sound plausible at first, we hold that it is mistaken. We now refute it by a counter-example based on a real-life case. We quote from (Murphy, 2017):

> on May 30 1842, John Francis shot at Queen Victoria, riding in her carriage outside Buckingham Palace on Constitution Hill. This was actually Francis's second attempt; the day before, he had pulled out his pistol but had either lost courage or his gun had misfired; he slipped away. But three people had witnessed him, and one of these was Prince Albert. The royal couple, then, was aware that an assailant was on the loose, and they thought it more than likely he would strike again.

Now imagine that on May 30th Prince Albert sees Francis, unaware that he has been spotted, loading his pistol, getting ready to make his attempt. Albert deliberately refrains from raising the alarm until Francis lifts his arm and points, whereupon Albert sounds the alarm, and Francis is caught, as intended, red-handed. In this case, Albert deliberately refrains from sounding the alarm, and does so with the intention that Francis perform an action, pointing a gun at the Queen, that we may take to be (and was taken by Albert to be) impermissible. It seems to us that Albert's act of refraining here, taken with his intention that Francis point the pistol at the Queen, is, contrary to the Purist Principle, permissible.

Our example is couched in terms of an omission, but it seems to us that the example could easily be changed to feature a positive action. Suppose, for example, that Francis starts to abort the assassination attempt as Prince Albert on his horse is blocking his light. Albert wishes that Francis be caught red-handed, and moves out of the light, intending that Francis load his pistol and point it at Victoria. Here Albert performs an action, moving out of the light, with the intention that Francis perform an impermissible action, pointing the pistol. Yet it seems to us that Albert's action, taken together with his intention that Francis perform the impermissible act, is permissible, contrary to the Purist Principle.

Our examples show, we take it, that the Purist Principle is untenable, and that, accordingly, the objection from intention is unsound. We started out from a target-centred

version of the objection from temptation, and we moved from it to increasingly more plausible, but, we have argued, ultimately unsound, agent-centred objections that were recoverable from it. The most plausible of these, namely the objection from intention, was based on the Purist Principle. Given that, we argued, even this objection was unsuccessful, is there any decent agent-centred objection in the vicinity? We believe that there is, although the one we are about to describe has a more modest goal, in that it is not designed to show that entrapment is impermissible. We call this new objection 'the objection from moral alliance'. We do not suggest that this objection is the only viable moral objection to entrapment. Neither do we suggest that targeted-centred considerations, such as the appeal to rights of the target that an act of entrapment may be held to breach, have no important role to play. Rather, our aim, in introducing the objection from moral alliance, is to draw attention to the interesting nature of this new objection and, especially, to its value in explaining why entrapment is harder morally to justify than intentional temptation and virtue testing.

5. THE OBJECTION FROM MORAL ALLIANCE

For parity with the arguments that we have criticized, let us set out the objection from moral alliance in standard form, before we proceed to explain it in more detail.

1.	In the cases of entrapment that are of interest to us (see Section 2), the agent procures the target's performance of an impermissible act.	Premise
2.	An agent that procures (or attempts to procure) the performance of an act thereby becomes more closely morally allied with that act than would have been the case if the agent had, other things being equal,	Premise

	merely intentionally tempted the target into the act's performance (that is, without attempting to procure it).	
3.	The more closely morally allied an agent's action is with a target's impermissible act then, other things being equal, the worse that agent's action is, morally speaking.	Premise
4.	An agent that procures (or attempts to procure) a target's performance of an impermissible act thereby does something morally worse than would have been the case if the agent had, other things being equal, merely intentionally tempted the target into the act's performance (that is, without attempting to procure it).	From 1, 2, 3

Entrapment to perform an impermissible action is morally objectionable, from an agent-centred point of view, because it allies the agent, via the agent's procurement (or attempted procurement) of it, with the impermissible act. The attempted procurement of the act involves the agent, on our account, and at least if it is a good attempt, in commending, requesting, or enjoining the performance of the target's impermissible act.

Suppose that each of two agents, *A* and *B*, intends that a target, *C* in the case of *A*, and *D* in the case of *B*, perform an impermissible act. *A* decides to entrap *C* by enjoining *C* to perform this act. *B* decides merely to present to *D* the opportunity to perform the same act, with the hope that *D* will perceive this opportunity as alluring, but without enjoining it or recommending it to *D* in any way. In short, *A* decides to entrap *C*, and *B* decides intentionally to tempt, but does not intend to entrap, *D*. Other things being equal, *A*'s act is the more objectionable. This is because in performing the communicative acts that are intended to procure *C*'s act, *A* becomes strongly *allied* with that act. This is similar to the common suggestion that lying is worse, *ceteris paribus*, than mere deception, even if both involve the

intention to produce in the other party a false belief: in lying, one allies oneself, by means of one's verbal act, more thoroughly with the false content than one does in mere deception.[14] Indeed, in lying one not only commits oneself to the false content, but usually intends that the target believe the false content precisely on the basis of one's committing oneself to it.[15] Of course, it is not merely in lying that one allies oneself with the content of what one says; on the contrary, this is a, perhaps the, distinguishing characteristic of assertion (cf. Brandom, 1983). The reason why one does not, in asserting a conditional, assert its antecedent and its consequent as well, is that one does not ally oneself with the antecedent and the consequent, just with the conditional as a whole.

An argument for this suggestion, that entrapment is harder morally to justify than mere temptation in virtue of the fact that in entrapment the agent is strongly allied with the impermissible act, can be found in consideration of the reaction of the target on realizing that they have been entrapped. The target may reasonably say 'but you asked/advised/told me to do it!'. It seems to us that this protest carries force. To see the contrast, suppose that the target has instead been *merely tempted* into performing the illegal or immoral act, i.e. in this case there has been no procurement of the illegal or immoral act. Here the protest 'but you *intentionally tempted* me to do it' does not seem to us to carry the same force, even if the target realized at the time that the agent also intended that the target perform the action. (To recur to our earlier

[14] The idea that liars ally themselves with the false content receives a very helpful discussion in (Timmermann and Viebahn, 2021). The idea is also discussed, though there described using the word 'warrant', in (Carson, 2018: 158–59).

[15] We are grateful to Fredrik Nyseth for this point. It is not always the case, however, that in lying one intends that the target believe the false content precisely on the basis of one's committing of oneself to it; sometimes when lying one intends only that the target believe that *one believe* the false content.

example, suppose that Francis had protested that Albert had intended that he point his pistol at Queen Victoria: such a protest would seem to us to carry little moral weight.)

So, the main agent-centred moral problem with entrapment, we suggest, has to do with the element of moral alliance that entrapment involves. Entrapment to perform an impermissible act is morally objectionable, from an agent-centred point of view, because it allies the agent, via the agent's procurement, or attempted procurement, of it, with an impermissible act. The attempted procurement of the act involves the agent, on our account, and at least if it is a good attempt, in commending, requesting or enjoining the performance of the target's impermissible act. This, we noted, was structurally similar to a common moral stance against lying: when one entraps, one allies oneself with the impermissible action that one procures, just as the liar is said to do with respect to the content of the lie. While the two stances are logically independent, just as this stance on lying holds that lying is harder for an agent morally to justify than is mere deception, we similarly hold that entrapment is harder for an agent morally to justify than is mere temptation.

One might think that it is permissible for an agent to counsel or recommend an impermissible action, but only if the agent intends that the target decline the invitation.[16] We think, however, that if the agent requests or enjoins the impermissible action, intending so to do, then that suffices to make the agent's action harder morally to justify than mere temptation would be, whether or not the agent intends that the request be refused.

Here is a difficult case. Suppose that a principled junior police officer, opposed to entrapment, is instructed by a superior to clear up the problem of drug pushers in a certain area by going undercover and requesting drugs from suspected pushers. Would it be a morally bad

[16] This seems structurally similar to the view that it is permissible intentionally to assert something one believes to be false, but only if one does not intend that the hearer believe the assertion. Kant (1997: 28) seems to hold this view; for discussion, see (Mahon, 2009: 207).

course of action for the principled junior officer to obey the letter of the superior's command by saying exactly the things usually said in cases of entrapment (i.e., saying 'may I buy some drugs?' etc.) with targets thought *unlikely* to yield to the request? It might be thought that doing so would yield the best of all possible worlds: the principled junior officer is happy that no entrapment has occurred, and keeps their job, and the senior officer, in blissful ignorance of the full picture, is happy that drugs have been requested undercover, albeit unsuccessfully. But we demur. We think that even in this case attempting to procure the impermissible action (even if not technically entrapment) is harder to justify morally than mere temptation would be, since the principled junior officer is still—albeit reluctantly—requesting something that the officer thinks should not be requested.

Let us illustrate how the objection from moral alliance helps explain some relevant moral subtleties. Suppose that there is a right *not* to be entrapped, but that the right is alienable: a person can consent to relinquish the right within a certain context. Take the example of entering into a contract of employment as a police officer. Miller, Blackler, and Alexandra (2006: 141) note that policing is an endeavour in which corruption is an ever-present and serious danger, and that 'testing' is among the methods used to detect and to deter it. Moreover, they suggest, 'detection and deterrence that might not be acceptable in some other professions' can be justified by the 'tendency to corruption' that is, they think, inherent in policing. They add (2006: 142) that the testing may involve (what they take to be) targeted or random entrapment. Given all these suppositions, would the fact that the target gave a general form of consent, in undertaking to be employed as a police officer under such conditions, suffice for the target to lose their right not to be entrapped? While we do not want to hold that one can 'contract away' all one's rights, we do think that in the case of entrapment it is possible and permissible for an individual voluntarily to give up their right not to be subject to entrapment (cf. Feinberg, 2003: 76), and we can easily imagine situations in which it would be permissible

for a law-enforcement agency to make it a condition of employment that employees 'contract away' their right not to be subject to entrapment. By the lights of the objection from moral alliance, however, the common view that entrapment should be a method of last resort is vindicated *even in cases where the target has generally consented to be entrapped*. Indeed, the objection from moral alliance is the only objection here considered that explains why, even in situations in which entrapment might be permissible, entrapment remains harder morally to justify, other things being equal, than mere intentional temptation. If it is feasible to monitor and promote probity among members of the police force via (mere) intentional temptation, virtue testing, or both, then these methods are, other things being equal, morally easier to justify than entrapment is. A different reason why, on our account, these methods are morally easier to justify is that, in the cases of entrapment that are of present concern, when a target is entrapped the target actually does something that the agent considers wrong (under a given normative system that the agent takes to be binding on the target). This point, however, cannot explain why a failed attempt at entrapment is harder morally to justify than a failed attempt at (mere) intentional temptation. The objection from moral alliance can explain this moral difference: according to the objection, in cases of entrapment, it is the communicative acts that figure in the agent's procurement, or attempted procurement, of the action that the agent intends that the target perform that morally ally the entrapping agent with it. These communicative acts ally the agent morally with the planned act even in cases in which the agent's attempt to entrap the target is unsuccessful, either because the agent failed to incline the target's will towards the planned act or for some other reason.

6. ON THE MORALITY OF VIRTUE TESTING AND MERE TEMPTATION

While we have made several passing remarks about the morality of virtue testing and mere temptation, we now go into some more detail. While it is clear from our previous discussion, given that these phenomena do not involve procurement, that the objection from moral alliance does not apply to them, the question of their ethical acceptability remains.

Take virtue testing first, and distinguish two general cases: when the agent intends that the target pass the test, and when the agent intends that the target fail the test. In the first case, we think that virtue testing is generally morally acceptable. It is sometimes acceptable to present someone with the opportunity to do wrong even if one does not consider them trustworthy. For example, one might hope against hope that they pass the test even if one does not think this likely.

Still, we do not mean to suggest that it is always acceptable for an agent to test a target if the agent intends that the target pass the test. It depends on what the risks and benefits are. If the agent knows that the test is relatively unimportant, but that the target's failure to pass the test—or even the stress of the target's being subjected to the test itself—will cause great harm and upset to the target and many others, then it will usually be unacceptable to subject the target to the test. Further, we do not believe that just any old motive for testing the target will suffice, even if, as it happens, there is a lot at stake. For example, if the agent tests the target just for a laugh, or to see the target squirm under pressure, that would not justify the test, even if it happens to be the case that the world would be a lot better as a result.

Take now the other general case of virtue testing, in which the agent intends that the target fail the test. As could be inferred from our counter-example to the Purist Principle, we hold that this is sometimes acceptable.

We hold, then, that entrapment, temptation, and virtue testing can be permissible, though we have said that entrapment is always harder to justify than are mere temptation and mere virtue testing. What features make entrapment, temptation, and virtue testing potentially bad actions, in need of justification? We hold that there are two principal bad-making or aggravating features: (i) if the agent believes that the target is not disposed to perform the impermissible action (or something as bad or worse) in the future, and (ii) if the agent does not intend that, as a result of the target's performing the impermissible action now, the impermissible action (or an equally bad or worse action) will not be performed later.[17] These features would, if not outweighed, suffice to make entrapment or temptation or virtue testing (as the case might be) impermissible.

Both condition (i) and condition (ii), and their conjunction, can be outweighed, however. For example, suppose that the Chief of Police of Town *X* believes that all his officers are honest, but, in order to prevent vast civil unrest, needs to be able to demonstrate this fact. To this end he leaves money lying around the police headquarters in Town *X* in order to weed out any subordinate officers that there may be that are likely to steal money. Here the Chief does not think that the officers are likely to steal in the future unless caught now, since he thinks that they are honest, so condition (i) holds. His motive in leaving the money lying around is

[17] It might be wondered how these conditions differ from others in the literature. For example, (Miller and Gordon, 2014: 276) argues that there are five conditions that, if jointly satisfied, render entrapment permissible. These conditions include that the target be disposed to perform the action in question (or an equally bad or worse action, we should add), and, indeed, they go further, stating that the target should have a 'standing intention' to perform it. Although another of the conditions rules out that the agent create the crime, this does not quite correspond to our (ii). Our (ii) states that it is a bad-making feature if the agent's intentions in tempting do not include the prevention of crime, if, for example, the agent did it solely to gain power over the target. We suspect that Miller and Gordon simply take for granted the intention, on the part of the agent, to prevent crime. We are grateful to an anonymous referee for this reference.

not in order to prevent stealing, but rather to demonstrate that Town *X*'s police force is honest. Indeed, the Chief might even think that any bad apples there might be would simply sign up to become officers of the police of Town *Y*, and steal even more under that lax regime. It follows that condition (ii) also holds. Nevertheless, it seems clear to us that his goal of demonstrating the honesty of the police force of Town *X* and thereby avoiding widespread civil unrest would justify his action, even if he had to engage in (mere) temptation to do it. Would it also justify the Chief in *entrapping* for this goal? As we said earlier, entrapment is harder to justify ethically than is mere virtue testing or mere temptation, but if the stakes were high enough (i.e. the possible consequences of the unrest in our example serious enough) we should not want to deny that the badness of entrapment could be outweighed.

7. CONCLUSION

We have distinguished three related phenomena: virtue testing, temptation, and entrapment. Making use of previous work, we provided a definition of entrapment, which introduced the key notion of procurement. Virtue testing is, in a sense, the weakest of the three notions: it need not involve either procurement or temptation. Temptation and entrapment differ principally in that the former need not involve procurement. While we hold that all temptation is virtue testing, we deny that all virtue testing is temptation; we deny that all temptation is entrapment, and we deny that all entrapment involves temptation.

We have considered two attempts to show that entrapment is wrong: the objection from temptation and the objection from intention. We have rejected each of them, and we have spent more time discussing the latter, more plausible objection. It is based on an initially attractive position, the Purist Principle, that it is impermissible to intend that someone perform an impermissible act. We have argued that this principle should be rejected. Instead, we have argued for what we called the 'objection from moral alliance'. According to it, entrapment is

morally worse than mere intentional temptation because, when one entraps, one strongly allies oneself with the wrongful action that one procures.

Lastly, we have argued that, while virtue testing and (mere) temptation are not subject to the argument from moral alliance, they, like entrapment, are not always acceptable. We have, making use of our work discussing the Purist Principle, identified two important bad-making conditions that, if present, need to be outweighed for the entrapment or temptation or virtue-testing to be permissible. We also gestured towards further considerations concerning risks, benefits, harms, and motives that could potentially outweigh these bad-making conditions. In short, our final position is that, while it is in many circumstances morally permissible to test a target, and in some circumstances permissible intentionally to tempt a target, the action of entrapping a target into doing something impermissible is, while not always impermissible, harder to justify morally than either of these.

What we have mainly sought to do here is to recover what we take to be ultimately salvageable from the objection from temptation, and it has turned out that this, in the form of the objection from moral alliance, is an agent-centred objection to entrapment rather than a target-centred one. Nevertheless, we intend our discussion to have been without prejudice to the question whether acts of entrapment are typically wrongs to, or breaches of the rights of, the target. We have not ruled out that they are such wrongs, either for a reason independent of the objection from temptation or because acts of entrapment typically involve morally unacceptable cases of intentional temptation that wrong the target. Nevertheless, the power of the objection from moral alliance, and some of its interest, lies in the fact that it serves to explain, on the one hand, why acts of entrapment are harder to justify morally, *even in such*

cases, than acts of mere temptation and of virtue testing and, on the other hand, why failed entrapment attempts are morally worse than failed attempts at temptation or virtue testing.[18]

REFERENCES

Altman, Andrew and Steven Lee. (1983) 'Legal Entrapment'. *Philosophy and Public Affairs*, 12 (1), 51–69.

Brandom, Robert P. (1983). 'Asserting'. *Noûs*, 17(4), 637–650.

Carson, Thomas L. (2018) 'The Range of Reasonable Views about the Morality of Lying'. In Eliot Michaelson and Andreas Stokke (eds), *Lying: Language, Knowledge, Ethics, and Politics* (Oxford: Oxford University Press), 145–160, https://doi.org/10.1093/oso/9780198743965.003.0008.

Dillof, Anthony M. (2004) 'Unravelling Unlawful Entrapment'. *Journal of Criminal Law and Criminology*, 94 (4), 827–896, https://doi.org/10.2307/3491412.

Dworkin, Gerald. (1985) 'The Serpent Beguiled me and I Did Eat: Entrapment and the Creation of Crime'. *Law and Philosophy*, 4 (1), 17–39.

Feinberg, Joel. (2003) *Problems at the Roots of Law: Essays in Legal and Political Theory*. New York: Oxford University Press.

Fletcher, George, P. (1978) *Rethinking Criminal Law*. Boston: Little Brown.

Foot, Philippa. (1967) 'The Problem of Abortion and the Doctrine of Double Effect'. *Oxford Review*, 5, 5–15. Reprinted in Philippa Foot, *Virtues and Vices, and Other Essays in Moral Philosophy* (2nd ed., Oxford: Blackwell, 2002), 19–32.

Ho, Hock Lai. (2011) 'State Entrapment'. *Legal Studies*, 31 (1), 71–95, https://doi.org/10.1111/j.1748-121X.2010.00176.x.

[18] ACKNOWLEDGEMENTS.

Hill, Daniel J., Stephen K. McLeod, and Attila Tanyi. (2018) 'The Concept of Entrapment'. *Criminal Law and Philosophy*, 12 (4), 539–554, https://doi.org/10.1007/s11572-017-9436-7.

Hughes, Paul M. (2004) 'What is Wrong with Entrapment?'. *Southern Journal of Philosophy*, 42 (1), 45–60, https://doi-org.liverpool.idm.oclc.org/10.1111/j.2041-6962.2004.tb00989.x.

Hughes, Paul M. (2006a) 'Temptation, Culpability and the Criminal Law'. *Journal of Social Philosophy*, 37 (2), 221–232, https://doi-org.liverpool.idm.oclc.org/10.1111/j.1467-9833.2006.00329.x.

Hughes, Paul M. (2006b) 'Temptation and Culpability in the Law of Duress and Entrapment'. *Criminal Law Quarterly*, 51 (3), 342–359, https://doi-org.liverpool.idm.oclc.org/10.1111/j.1467-9833.2006.00329.x.

Kant, Immanuel. (1997) *Lectures on Ethics*, ed. Peter Heath and J. B. Schneewind, tr. Peter Heath. Cambridge: Cambridge University Press.

Mahon, James Edwin. (2009) 'The Truth about Kant on Lies'. In Clancy Martin (ed.), *The Philosophy of Deception* (Oxford: Oxford University Press, 2009), 201–224, https://doi.org/10.1093/acprof:oso/9780195327939.003.0012.

Miller, Seamus and John Blackler. (2005) *Ethical Issues in Policing*. Aldershot: Ashgate.

Miller, Seamus, John Blackler and Andrew Alexandra. (2006) *Police Ethics*. 2nd ed., Winchester: Waterside Press.

Miller, Seamus and Ian A. Gordon. (2014) *Investigative Ethics: Ethics for Police Detectives and Criminal Investigators*. Chichester: John Wiley and Sons.

Murphy, Paul T. (2017) 'Shooting Victoria' blog, 29 May 2017. Retrieved 20 December 2020, from https://shootingvictoria.com/post/161208663221/deja-vu-all-over-again.

Park, Roger. (1976) 'The Entrapment Controversy'. *Minnesota Law Review*, 60, 163–274.

Smith, Russell E. (1995) ‘Formal and Material Cooperation’. *Ethics and Medics*, 20 (6), 1–2.

Stitt, B. Grant and Gene G. James. (1984) ‘Entrapment and the Entrapment Defense: Dilemmas for a Democratic Society’. *Law and Philosophy*, 3 (1), 111–132.

Timmerman, Felix and Emanuel Viebahn. (2021) ‘To Lie or to Mislead?’. *Philosophical Studies*, 178, 1481–1501, https://doi.org/10.1007/s11098-020-01492-1.